CHAPTER 12

DATA SOURCES AND STATISTICAL ISSUES

INTRODUCTION

The foundation of cost and return (CAR) estimation is the data on which those estimates are based. Therefore a critical step toward establishing uniformity in the methodology used in generating CAR estimates is to examine the sources of these data and to investigate conditions under which each may provide a suitable basis for analysis. The purpose of this chapter is to encourage the analyst to look closely at the data, their strengths and weaknesses, and their suitability in the specific context in which the analyst is producing CAR estimates.

This chapter discusses and compares the most common sources of data for CAR estimates. Data for CAR studies can be obtained in a variety of ways including the use of large-scale probability surveys designed to collect primary data about cost of production, the use of data from farm records systems, the use of information obtained from a single farming operation, and the use of agricultural engineering equations based on field data. Each has its place in providing data for CAR estimation. The chapter also examines a variety of statistical issues that are relevant to obtaining and using data in analysis and estimation.

House, in her remarks at a conference on CAR estimation, stated that quality data must provide estimates that are "accurate, defensible, affordable, and ... target the desired population" (House: 81). There are many factors in the data collection process that are important for assuring those qualities, but perhaps the most important is statistical inference. Statistical inference determines whether, and to what extent, results from the analysis and estimation can be generalized to a broader set of farming operations. Statistical inference is largely determined by two activities: precisely defining the group (or target population) the analyst wants to investigate, project for, or draw conclusions about; and selecting representative data from that population for the analysis.

Defining the Target Population

In the context of this publication, the target population for a data collection activity is the group about which the analyst wishes to make CAR estimates. Commonly the target population will be the subset of farms engaged in a very specific farming enterprise within a localized geographic area. Examples of target populations in this context are all farms engaged in dairy production in the upper Midwest, or farms engaged in cotton/almond production in the San Joaquin Valley. For certain purposes the target population may need to be defined even more precisely. Extension economists may want to develop CAR estimates that are representative of progressive, well-managed farms (rather than all farms) engaged in the selected enterprise because those estimates may be more useful in guiding potential producers. On the other hand, the United States Department of Agriculture (USDA) and others producing historical estimates will generally want to include a broader geographic area and to target all farms engaged in the enterprise regardless of whether they are progressive or not.

Chapter 12. Data Sources and Statistical Issues

Many different groups can be targeted legitimately for CAR estimation depending on the analytical needs and budget of the particular endeavor. A problem arises, however, when an analyst targets one population for CAR estimation but chooses a data collection method that actually focuses on some other group.

Selecting Representative Data

The second step in the process of obtaining data is to select a representative sample of data from the target population in such a way that valid inferences can be maintained. Two general types of samples are possible: a statistical sample, or a judgment sample.

A statistical (or probability) sample is one in which each farm in the targeted population has a positive and knowable chance of being included. The probabilities of inclusion in the sample are used to produce sample weights, which in turn convert the estimates produced from the data into estimates representative of the entire target population. Statistical theory then helps the analyst describe certain measures of the accuracy of these estimates. Because it gives a chance of selection to every farm in the targeted population, and because it can provide measures of accuracy, a statistical sample is considered to be superior to a judgment sample. If one can obtain data for CAR estimation from a statistical sample, that may be the best procedure. However, statistical sampling procedures can be very expensive, they don't produce accurate results for very small sample sizes, and they are subject to a variety of other types of collection errors. Thus, there are many situations in CAR estimation where carefully selected judgment samples are appropriate vehicles for obtaining data.

A judgment sample is selected from the population through some method other than statistical sampling, usually the subjective decision of one or more individuals. This means that at least some units in the population have no chance of selection and/or it is not possible to determine what the selection probabilities (and thus the sample weights) are. Williams indicates that judgment samples are problematic because "the accuracy of judgement samples cannot usually be determined. They are not necessarily inaccurate, but if they are accurate the accuracy is usually unknown and depends upon the expertise of the specific individual [selecting the sample]" (Williams: 47). In general there can be little empirical assessment of the accuracy of CAR estimates made from data collected from a judgment sample. The analyst should be aware of this serious limitation before choosing this approach to data collection.

DATA SOURCES

This section is divided into two parts. The first part presents a brief discussion of three alternatives for generating data that can be used to produce CAR estimates and/or to analyze various other aspects of the structure of production in farm firms. The second part consists of a brief review of several studies that have compared the implications of using different data sources from the same general population to examine CARs and other farm level characteristics.

Alternatives for Generating Data

Although there are numerous alternatives for generating data for CAR estimates, three of the most commonly used methods will be discussed here. The three alternatives for generating data required in the preparation of CAR estimates for agricultural commodities are (a) probability surveys, (b) farm record systems, and (c) the economic engineering approach.

Probability Surveys

The major source of data collected using probability surveys is the federal government. Three sources of farm survey data identified by J. D. Johnson are (1) the Census of Agriculture, (2) special follow-on surveys to the Census of Agriculture (e.g., farm finance or irrigation surveys), and (3) USDA farm economic surveys, particularly the Agricultural Resource Management Study (ARMS). This comprehensive set of surveys was previously conducted under the name Farm Costs and Returns Survey (FCRS). A second source of data generated from probability surveys is work that takes place in a few states under the auspices of land grant universities.

For the purposes of the present discussion, we want to focus on two specific probability surveys, one at the federal level and the other at the state level. The federal-level survey is the Agricultural Resource Management Study data collection process, a cooperative project by USDA's Economic Research Service and National Agricultural Statistics Service. The state-level discussion centers on procedures used by the Louisiana and Mississippi Agricultural Experiment Stations.

The ARMS data are collected through a nationwide survey of approximately 26,000 farm and ranch operators. A main objective of the survey is to collect data to develop weighted average costs of production (COP) estimates for specific farm commodities. Each observation is intended to be representative of a number of similar farms. In addition, the survey generates data that are useful in examining a variety of farm-level issues such as efficiency, income and wealth levels, capital formation, and financial structure (J. D. Johnson). The commodities included in the survey depend on legislative mandate and on USDA needs. Given the high costs associated with conducting the surveys, data are collected for individual commodities every four to five years on a rotating basis (Morehart, Johnson, and Shapouri).[1]

Activities in Louisiana and Mississippi illustrate the generation of data from probability surveys at the state level. Although they work independently, the Experiment Stations in these two states follow very similar procedures to produce data for CAR estimation. Both stations cooperate with the National Agricultural Statistics Service (NASS) to conduct probability surveys of producers of major enterprises within various regions of the states. The surveys are conducted for each enterprise on a rotating basis every three years. The data collected from the surveys of producers are used to identify farming practices and the type and quantity of materials typically used by the farms in the various regions. The producer surveys are supplemented by

[1] For details concerning the ARMS data see J. D. Johnson, and Morehart, Johnson, and Shapouri.

state-wide surveys of suppliers to collect input price information and by additional information provided by extension personnel.[2]

When compared to other major sources of data for CAR estimation, the advantages of probability surveys rest in two important areas. First, only a probability survey can provide statistical inference to a broad array of potential target populations within the farm sector. Second, one can statistically describe the accuracy of CAR estimates that were based on probability data. With other sources, the accuracy is hard to measure and generally unknown. These points are discussed by House. However, data collection costs pose a major disadvantage to using probability surveys to obtain data. Another disadvantage is that few if any of these probability-based data sets are collected longitudinally. Data are typically collected once for a cross section with no follow-on in subsequent years. Hence, these data sets lack the richness over time often available through methods discussed in the following sections. Furthermore, problems outlined in the section entitled "Reliability Issues With Data" are applicable to data collected through probability surveys. Finally, probability surveys are likely to have confidentiality restrictions imposed upon release of their data, making these data sets difficult to share with the research community.

Farm Record Systems

Stovall and Hoover indicate that farm record keeping systems emerged during the 1940s in several midwestern states and gradually spread to other areas primarily in response to the increasing complexity of income tax rules. The data for these systems are provided by farmers on a voluntary basis and are used to prepare farm business analyses which in turn are used by farmers for management decisions. In some cases, the business record associations prepare the farmers' tax returns.

The record systems are usually sponsored by farm management associations, departments of agricultural economics, Cooperative Extension, vocational-technical schools, and/or Farm Credit Banks. According to Casler, at least two dozen states currently have some type of farm record system. Traditionally, the main use of the data generated through farm record systems has been for extension and education programs. More recently, however, several researchers have utilized these data to examine a wide range of issues in production economics and agricultural policy.

Two types of data result from farm record systems: (1) original individual farm records, and (2) aggregate business summaries that are typically published annually by the various systems. These summaries often subdivide the farms by size class, location, and/or some other salient feature and provide detailed physical and economic data for each group.

A major advantage of the data stemming from farm record systems is their high level of accuracy because of the scrutiny usually given to the information by field supervisors (Batte and Sonka). Another

---

[2] Empirical studies using data generated from probability samples can be divided into econometric and noneconometric. Examples of econometric studies are the papers by Weaver and Lass, and Huy et al. Examples of noneconometric studies are the papers by Hazilla and Kopp, Cooke and Sundquist, and USDA/ERS 1990a, 1990b.

important advantage is that in many cases individual records are available for the same farmers over a number of years, which makes it possible to apply panel data analysis techniques. In addition, the annual summaries are a unique source of high-quality, disaggregated time series data, which researchers have just begun to exploit. Unfortunately, these data also present several problems. One problem is the difficulty in building a data set covering several states. Even if access to data is secured from several states, the procedures used across states are often incompatible (Casler). Another important shortcoming of business records is that the farms included in these data sets do not proportionately represent the entire farm population. Willimack estimated that only 11% of farming operations across the country currently use such a service. Furthermore the percentage differs by type and size of farm. Thus, inferences to sizable portions of the farm sector are problematic, and no conclusive statements can be made concerning those populations. The extent to which data from business records differ from that of random samples for a given population has received some attention in the literature and is a point which will be discussed later.[3]

The Economic Engineering Approach

The economic engineering approach has been widely used to generate data required to produce CAR estimates, as well as estimates of other features of farm operations, particularly economies of size. This approach combines input-output information gathered from engineering, biological, and other relevant technical disciplines with information collected from the field (e.g., from farmers and extension agents) and with accounting data to estimate CARs or other measures of performance. Developing CAR data via economic engineering often requires that the researcher define a typical farm situation.

The procedures used to generate data, define variables, and characterize the typical or representative farm can range from very formal to very casual. The more formal procedures make use of the Delphi method (Pill). The methodology used in the more casual cases varies greatly from case to case (Klonsky 1992). Casual procedures tend to be the most frequently used in the field.

In order to illustrate a formal procedure to generate data for representative farm analyses, consider the approach used by researchers from the Agricultural & Food Policy Center (AFPC) located in the Department of Agricultural Economics at Texas A&M University. For several years, researchers at the AFPC have used whole-farm simulation models to examine the effects of farm programs on representative agricultural firms for various regions of the United States. The information required to construct the farm models is collected from producer panels for a particular type of operation in a given region. The producer panels provide information on the size of the typical operation, tenure arrangements, enterprises, costs of production for the various enterprises, crop yields (expected and historical), and machinery complements (Knutson et al. 1992). After the information is collected and processed it is reviewed by the panel members. The data are then incorporated into the farm-level policy model to produce pro forma financial statements for the panel farm. These statements are once again reviewed, adjustments are made, and this process is repeated until the panelists are

---

[3]Empirical studies using individual records have been published by Batte and Sonka; Hornbaker, Dixon and Sonka; and Bravo-Ureta and Rieger, among others. Examples of studies relying on annual summaries of business records are the papers by Adelaja; Foster and Rausser; Quiroga and Bravo-Ureta; and Cocchi.

satisfied that the financial projections are reasonable. Additional data are collected for each region with the assistance of appropriate land grant personnel. Information collected in this fashion includes interest rates, Commodity Credit Corporation (CCC) loan rates, prices received for outputs and paid for inputs, and income tax information. Finally, macroeconomic data, policy assumptions, and prices for policy analyses are obtained from the Food and Agricultural Policy Research Institute (FAPRI) located at the University of Missouri-Columbia and Iowa State University (Knutson et al. 1992).

The economic engineering approach is particularly useful in examining a priori the impact of possible changes in a wide range of variables such as technology, government programs, yields, and prices. This type of data, however, does not provide information on the actual farm situation. It simply illustrates what the situation would be if the assumptions incorporated into the analysis were to materialize.[4]

Comparisons of Estimates from Alternative Data Sources

The first attempt to compare formally the characteristics of farmers participating in farm record systems with those of farmers selected in random samples was undertaken by Hopkins in a study published in 1939. Hopkins compared record keeping farms with farms from a random sample and found that the managerial capacity of the former group was "definitely superior" to that of the latter. He concluded that "operating on land of approximately equal value and directing equal amounts of labor, but utilizing more short-lived capital, the record-farmers have obtained a significantly greater output and have earned higher net incomes" (Hopkins: 276).

In another early study, Mueller compared data collected by the Farm Bureau Farm Management Services for 210 cooperators from six counties in western Illinois with 193 farmers from a random sample of the same six counties. Mueller found that the record keeping farms, compared to their survey counterparts, were larger in size, had a higher investment per acre, were located on higher-quality soils, and had higher output per acre for the major grains. Land use patterns, however, were very similar in the two groups. Nevertheless, Mueller found no evidence of differences in the managerial ability of the two groups of farms. This author concluded "that differences between record keeping farms and a representative sample of all farms are essentially differences in the quantity of basic resources, particularly land and capital, utilized by the farm operators" (Mueller: 292).

Olson and Tvedt contrasted annual farm averages from the Southwestern Minnesota Farm Business Management Association (FBMA) with U.S. Agricultural Census averages for 1982. The authors' specific objective was to examine the proposition that producers belonging to farm management associations are better managers and have larger operations than the population of farmers as a whole. The authors found that the average farm belonging to the association was larger and had higher crop yields than the average census operation. In addition, the authors found that association farms had higher total investment and production expenses, which is consistent with their larger size. On a per acre basis, however, investment and production

[4]For a comprehensive, although dated, review of economic engineering studies see Madden. A more recent review was published by H. R. Jensen.

expenses were lower for association farms, which might be a reflection of higher efficiency and/or better land quality.

In a later study, Andersson and Olson examined 1987 Southeastern and Southwestern Minnesota FBMA and FCRS farm record data. They were able to use individual farm observations and thus provide a more accurate comparison to FCRS estimates than the previous Minnesota study. The specific objectives of Andersson and Olson were to examine statistically any differences in farm characteristics between the two data sets, and to ascertain the farm size classes for which FBMA farms are statistically representative of those in the FCRS system. These objectives were pursued through descriptive and statistical comparisons of several variables reflecting a variety of farm characteristics.

Andersson and Olson found that the FBMA farms are not representative of the population in the study areas. They found major differences in overall farm size, number of tillable acres, rented land, and livestock production (particularly hogs). The differences in these variables led to marked divergences in farm income between the two farm groups, although solvency conditions were very similar.

Analysis focusing on farms with sales exceeding $60,000 still showed that FBMA operations had "a higher level of livestock production and a slightly larger tillable acreage mainly due to renting additional land. Economic performance measured by net farm income and returns to total assets and family labor was significantly...better for FBMA farms. So even though differences in...solvency positions were insignificant, the economic performance measured of the FBMA farms appears to be better than FCRS farms even in larger sizes" (Andersson and Olson: 310). Based on these results, the authors concluded that the FBMA data is not representative of all farms nor of all commercial farms.

Libbin and Torell set out to compare CAR estimates developed by researchers at New Mexico State University (NMSU) with USDA's CAR estimates for New Mexico farms and ranches prepared with FCRS data. The authors also compared estimates from crop CAR estimates developed in Illinois, Kentucky, and Missouri, and from livestock CAR estimates completed in Colorado and Washington, with figures published by the USDA. These authors found substantial differences in both crop and livestock CAR estimates for New Mexico. "Crop budget comparisons for selected states other than New Mexico yielded similar disparities in budget results. Livestock budgets from the two budget sources were similar" (Libbin and Torell: 308).

Koenigstein and Lins contrasted information obtained from farmers participating in the Illinois Farm Business Farm Management (FBFM) Association with FCRS farms for the year 1986. The Illinois FBFM Association was started over 40 years ago and has more than 7,000 farms, making it one of the oldest and largest associations of this type in the country. The authors used descriptive statistics to summarize individual farm records and focused primarily on financial variables.

The Illinois study found that about one-half of the FCRS farms were small, part-time operations, while the FBFM farms were larger and full-time. According to the authors, these size differences make direct comparisons across the two groups of farms difficult; however, farms that are in a similar size class had many characteristics in common. Koenigstein and Lins found that, because of omissions of several balance sheet items, the USDA estimated only 87% of the true net worth of the Illinois operations. Similar omissions of income statement items led the USDA to measure only 81% of the true net farm income.

Gustafson et al. performed statistical comparisons of financial characteristics for 1986 farm-level data obtained from the North Dakota Farm Business Management Education (FBME) program and from the USDA's FCRS. The former data source is based on 496 farms while the latter consists of 307 observations representing a total of 24,472 North Dakota farms. This study showed that the farms participating in the record keeping program had considerably more land, hired labor, production expenses, gross income, assets, and liabilities than their FCRS counterparts. In addition, the results revealed that "equity levels on record keeping farms are higher but profitability and returns to that equity are substantially lower" (Gustason et al.: 172).

The final study reviewed here is a detailed analysis of the costs of producing rice in Texas published by Rister et al. The Texas study used the AFPC farmer panel data approach discussed previously to produce CAR estimates for rice for four representative farming situations in 1989. A comparison of the per cwt cost estimates made by Rister and his colleagues with figures developed by ERS shows that in the former case the range is from $11.68 to $14.35, while for the latter the range is from $7.00 to $8.00. Rister et al. conclude their study by suggesting reasons contributing to this divergence in cost estimates. Among the most important was Texas rice producers' seeming inability to respond fully to the questions posed by the ERS survey (both because of question misunderstandings and due to producers' lack of time as a result of the questionnaire being administered during peak planting time in March and April), thereby underestimating expenses. Another difference was ERS's use of imputed returns to estimate several costs items that have traditionally had high government payment receipts. If these receipts are not included in the imputed returns, the costs associated with these assets are underestimated to the extent that these receipts have been capitalized into asset values. Some difference arose because ERS indexed variable costs between survey years but used actual yields for each year so that the cost of the higher-yielding semidwarf varieties released and adopted between survey years was not fully represented. Another difference related to the allocation of farm overhead expenses to planted rice acres versus all farm acres. As discussed in Chapter 9, this is an area where there is little hard data and allocations are often arbitrary. Another difference is that Rister et al. included expenses for drying, storage, marketing, and checkoff expenses that are required to transfer the crop to the first off-farm handler whereas ERS considered direct production costs only. The differences here are a good example of the difficulty in preparing costs of production estimates, and the importance of documenting assumptions so that users can adjust the estimates to fit a particular need or comparison.

RELIABILITY ISSUES WITH DATA

When one selects a sample (either a statistical sample or a judgment sample) from a population and uses information from the sample to represent that population, it is important that the representation be accurate. However, for reliability, it is equally important to know how accurate or inaccurate that representation is likely to be. Good data will be unused if the analyst does not consider them creditable. The situation is even more serious when inaccurate representations from data are perceived as correct and used as such. Without knowing the "truth," accuracy is difficult to measure. However, certain measurements and controls are possible, and this section will discuss some of these in more detail. Errors that reduce accuracy are usually categorized into two groups: sample variability and bias. To be accurate, data must allow estimation with low sampling variability and small biases.

Sampling Variability

Precision, one component of accuracy, measures how closely the results from a single sample are likely to match the results of a census conducted using similar procedures. (It does not tell whether those procedures are good or not!) An estimate is said to be "precise" when its sampling variability is small. With probability sampling, two positive results are possible. First, one can obtain objective and accurate measures of precision from the sample itself. Second, one can improve precision by simply increasing the sample size. When judgment sampling is used, one must look for subjective methods of measuring and increasing precision.

In probability sampling the standard error of an estimate is the basic measure of precision. The smaller the standard error, the higher the precision. A normalized form of the standard error, called the coefficient of variation (CV), is commonly used by analysts. The CV, given in percent, is the standard error of an estimate divided by the estimate itself. Confidence intervals can be computed using either the standard error or the CV, depending on whether one wants the interval expressed as an actual width or in relative percentage terms. A 95% confidence interval defines an interval around the estimate such that if the sampling procedures were repeated 100 times, the true (census) value would be within the interval for approximately 95 of the 100 repetitions. A 95% confidence interval has the width of four standard errors (or four CVs) and is centered around the estimate. The more "confidence" you demand in the interval, the larger that interval will be; for example, a 95% confidence interval is wider than a 90% confidence interval. If one were collecting data on the cost per hour for hired labor on farms in Illinois, for example, and decided to survey 500 farms using an equally weighted probability sample, the mean of this data divided by its standard deviation would be a measure of sampling variability. If this CV turned out to be 5, one would be concerned about using these estimates as representative of the cost of labor for all farms.

Bias and Its Sources

Confidence intervals can tell a lot about the accuracy of data when the data collection process is free of other types of errors. However, when things go wrong in the data collection process, one ends up measuring something different than what was intended. For example, an analyst may want to have an estimate of the total corn stocks held by an operation. Depending on how the question is asked, the farmer may only report those stocks physically on his property. The data collector has measured something different than what was intended. The bias is the difference between the value that was measured and the value one intended to measure.

Errors that can lead to such biases are referred to as nonsampling errors and they are universally hard to detect, measure, and control. This section will discuss three general types of nonsampling errors: response errors, nonresponse errors, and coverage problems. The focus will be on examples of such errors that are likely to occur in collecting CAR data, with discussions on techniques for avoiding or minimizing these errors.

Response Bias

Response error occurs when a respondent attempts to provide accurate information but fails to do so. The data collector is more often to blame than the respondent in these situations. Ambiguous and poorly worded questions are major causes of response error. Another common cause is asking the respondent to recall information from an earlier time period. Both situations are discussed here.

Question Design. Sudman and Bradburn give a simple principle in designing questions so that they can be answered accurately: use words that everyone will understand and that have only the meaning that is intended. Cost of production concepts are so complex that just getting agreement and consistency between analysts is difficult. It is small wonder that farm operators may have trouble understanding what is being asked. Garcia and Sonka point to how farm record systems obtain accurate information by following the principle extolled by Sudman and Bradburn: "A standard account record is used by all members. The accounting procedures used are documented both to the farmers and to the researcher. Farmers often work directly with field agents, reinforcing the standardization and the accuracy of the records" (Garcia and Sonka: 132). House points out that producer panels present a unique opportunity for proper interpretation of questions to respondents by allowing panel members to interact with one another as well as with the moderator to make sure that everyone is interpreting the questions in the same way. A comprehensive knowledge of the technical aspects of the enterprise (or firm) being analyzed is essential for developing a thorough CAR questionnaire and in administering a survey instrument. Thorough and consistent training of interviewers on the concepts being analyzed and on probing skills can go a long way to reducing the number of response errors during data collection, particularly those that relate to aspects of the operation that do not fit neatly into the designed instrument. Questionnaires sent by mail lose this aspect of quality supplied by the interviewer and therefore may be subject to more variation in the way respondents interpret questions. Dillman, however, makes a case for data collection by mailing questionnaires, and his book provides many helpful insights on how to instill quality in the design of mail questionnaires and how to achieve a reasonable response rate.

Recall Error. A second major type of response error results from the respondent's inability to recall information accurately. Two of the biggest recall problems are those of omission (forgetting to include certain items) and telescoping (including items outside the survey reference period). Omission is less likely to occur if the farm operator is asked for a list of specific expense items rather than a general question grouping types of expenses. A farmer reporting all feed expenses may omit expenses for supplements unless that is specifically listed. An appropriate time period is important if one is to minimize telescoping. In telescoping the respondent recalls and reports an expense that occurred outside the time period that the questioner intends. A farmer recalling hours of labor used during the previous four weeks may also include labor activity that occurred in the few days preceding that period, particularly if the earlier week was an active labor week. The best solution to both of these problems is to retrieve data directly from the farmer's own records. Clearly, this is the concept behind using farm records systems as a source of data for CAR analysis. However, record keeping is inconsistent among farmers and differs by type and size of farm. Johnson et al. report that approximately 70% of respondents on the FCRS used their records to answer questions on that survey. Twenty-three percent of respondents reported that they did not keep formal records of any kind.

Nonresponse Bias

When there is significant nonresponse on a data collection effort, the effect can be two-fold. First, it reduces the number of responses available for analysis, and thus reduces precision. The data collector should prepare for this situation in advance and increase the original sample size accordingly to assure that an adequate number of responses are available for analysis. Second, and more seriously, a nonresponse bias may occur if a correlation exists between the ability to get a response and the measurement of the item itself. In other words, the nonrespondents have certain characteristics different from those of the respondents, and those differences are important to the measurement of enterprise CARs. For example, if large farmers are less likely to respond in a data collection activity, the estimates from that data will be biased toward small farms and will underestimate farm size and any other variables closely related to farm size. The biases can be very serious.

Several steps are necessary to measure and control nonresponse bias. First, it is important to identify and classify reasons for nonresponse and the characteristics of these nonrespondents. Using that information, the data collector should modify data collection procedures to make them less burdensome to potential respondents and/or communicate more convincingly the reasons potential respondents should participate. The analyst should use that information to adjust rates so that the respondents more accurately represent the entire population. This concept is discussed in more detail in the section on "Reliability Issues With Analysis."

An example of this multistep approach to measuring and controlling nonresponse is the procedure used in conjunction with the FCRS (now ARMS). A special project was conducted in 1991 to identify and classify the reasons for nonresponse on this survey. The results, summarized by O'Connor, indicated that the single most frequently reported reason for refusing to respond to the survey was that the farmer was "too busy." Other reasons frequently given were that the "information requested was too personal," "that the farmer didn't like surveys," that his or her "farm records were at the tax advisors," or that "surveys and reports hurt the farmer more than help." In response to this information, data collectors began the development and testing of a shortened version of the questionnaire and developed materials to prepare interviewers to discuss concerns and grievances brought up by farmers. A separate analysis by Dillard, and then by Rutz and Cadwallader, indicated that response on the FCRS is correlated with both size and type of farm. In particular, they found that larger operators were less likely to respond on the survey. Turner and Burt quantified the size of the nonresponse bias and stated that total expenses were underestimated by approximately 10%. They began developing a procedure to group both respondents and nonrespondents by size and type classification groups, and to adjust the survey weights within those groups to allow respondents to represent nonrespondents within their classification grouping only. These procedures were successful in eliminating most of the identified bias.

Nonresponse bias exists in most data collection activities. It is important to understand that contacting only those producers likely to provide information (the approach often used in setting up producer panels or using economic engineering techniques) is, in fact, the same as excluding those unlikely to respond. Thus the potential for bias resulting from excluding producers reluctant to participate in surveys, producer panels, or record keeping services affects most data collection approaches for commodity CARs. Efforts must be made to examine the results for potential bias and adjust the estimates when necessary.

Coverage Bias

Earlier we discussed the importance of the data collector carefully defining the group or population which is to be targeted and choosing sampling and data collection procedures to assure that inferences can be made for exactly that group. When the sampled population is different from the target population, we say that coverage bias can occur. For example, a coverage problem occurs when data from a farm record system is used to make inferences about dairy farms. Willimack estimates that only 20% of dairy farms use farm record systems, and that these systems are used more often by larger farms. The question is whether that 20% can accurately represent the missing 80%.

Coverage problems may be moderated by reweighting techniques that adjust data so that it is more closely representative of the targeted population. Nonresponse, as discussed previously, can be viewed as a coverage problem. The group of respondents cannot provide inferences for the target population, which includes both respondents and nonrespondents. The solution suggested by Turner and Burt adjusts the survey weights based on classification groupings and auxiliary information. See "Appropriate Use of Weights" in the next section for more discussion.

RELIABILITY ISSUES WITH ANALYSIS

Two important issues involving analysis of data for CAR estimates are discussed in this section. The first is the appropriate use of weights in estimation and modeling to assure that the resulting estimates are representative of the enterprise being targeted. The second is the process of mixing data from different sources together to produce a single set of estimates. Beyond the material covered in this section, the appendix to this chapter contains two sections that are pertinent to the issues of data analysis. Appendix 12A provides an Overview of Statistical Sampling Techniques that can help an analyst understand the sample design which produced an existing data set. Appendix 12B provides guidelines for the appropriate rounding of estimates based on the accuracy of input data.

Appropriate Use of Weights

Weights are used with data obtained from individual farming units so that collectively those units are representative of the target population. As discussed earlier, that target population may be the entire farm sector or some designated portion of it. Weights are an essential component of data analysis regardless of whether the data come from a probability sample, a judgmental sample, or an economic engineering approach to collecting data. If weights are not used with the data, one is making the implicit assumption that each response is equally representative of the target population.

When data come from a probability survey, the weighted sum of the sample data generally will provide an estimate of the total for the item being measured. In these cases the weight is called an **expansion weight**. The survey design dictates the value of the weights. In complex designs there are different weights for each stratum and sampling stage. The weight is the inverse of the probability of selection of each unit and may be modified by poststratification and nonresponse adjustments. Expansion weights, summed by themselves across the sample, will equal the population total. **Relative weights** are often used in data analysis instead of expansion weights. Relative weights are calculated by dividing each expansion weight by the sum of all expansion weights (population total). Relative weights, summed by themselves, will always equal one. The weighted sum of the sample data, using relative weights, will provide an estimate of the mean for the item being measured.

Cost of production data frequently is produced without weights. The analyst must then decide whether he or she must develop weights to use in the analysis of this data, or whether unweighted results will be reasonable for the purpose at hand. The crucial judgment is whether each response (and there may be only one in a given data set) is equally representative of the farming enterprise being examined. If this is the case, unweighted analysis of the data set is appropriate. If not, weights must be developed.

The following is a simple example of how one could develop weights for analysis. Assume an analyst has access to cost of production data for dairy farms on a farm record system. Because data from such systems generally do not represent a cross section of the dairy farm population, the analyst needs to analyze the coverage of the data and make several judgment decisions. The data is broken into subsets of fairly homogeneous farms characterized by size, a technology index, and geographic location. Using census data supplemented by a university study quantifying the use of various production technologies within the state, the analyst produces estimates of the number of farms within the state that could be classified into each group. The estimates are constructed so that they will sum to the total number of dairy farms within the state. With these estimates, the analyst returns to the farm record data set. If the purpose is to develop cost of production estimates for each of these subgroups, the data within each subgroup can be used unweighted in the analysis. However, if there is a need to produce CAR estimates across subgroups, then the population counts developed to weight each subgroup must be used in a way that accurately reflects its relative size in state dairy production.

There are a variety of computer software packages available for analyzing weighted survey data. Most Statistical Analysis System (SAS) procedures have built-in options for using data weights. SUDAAN, available from the Research Triangle Institute, and PC CARP, developed at Iowa State University, provide options for analyzing complex survey data. Lee et al. provide a more extensive review of available computer software.

Mixing Data From Multiple Sources

It is very unusual for an analyst to derive CAR estimates from a single data source. Often there is a primary source with several secondary sources of data. The secondary data is sometimes used to produce estimates for selected activities within the enterprise. Other times the secondary data is used to develop weights for the primary data, so that the resulting CAR estimates are more representative of the target population. Census data are frequently used for this latter purpose. In particular, the economic engineering approach to developing CAR estimates generally uses a multitude of data sources to build a typical farm scenario.

Mixing different sources of data proves to be a very cost effective approach to CAR estimation. In many cases it is the only reasonable alternative. It is very important, however, to evaluate each source of data critically, both in terms of its overall quality and in terms of its compatibility with other data sources being used. Earlier parts of this chapter discussed data quality and reliability issues. Each data source should be evaluated independently in light of these issues. The composite of data sources can be no more reliable than its weakest member.

This section addresses data compatibility. Sometimes subtle differences in data sources can be important. Consider the following example. The primary data used by an analyst for CAR estimation has cost data for chemical inputs, but those costs are not broken out by individual chemical. Since a breakout is needed, the analyst gets recommendations for the use of various chemicals from an extension agent and price information from a chemical distributor. The analyst wishes to combine data from these three different sources to produce a cost breakout of chemical usage. The analyst must evaluate the inference that can be made from each data source and see if they are compatible with each other and with the overall target population about which the CAR estimates will apply. As discussed earlier in this section, the primary data must be representative of the type of enterprise targeted for the CAR estimates. Second, the recommendations from the extension agent must also be geared specifically to that same targeted enterprise, and not general recommendations for a certain commodity. Third, the price data must reflect the price structure most likely encountered by the targeted farmers. Germane to this issue would be geographic location, type of supplier, and quantity discounts. Finally, the combination of information must make sense together. The cost breakouts developed from the two secondary sources of data, when added together across different chemicals, must be consistent with the aggregates obtained from the primary data. If not, the analyst must search for the source of incompatibility before proceeding.

The above example stresses that different sources of data should target similar enterprises. Alternatively, use of secondary data for weights provides an example of when it is appropriate to utilize data sources that target a different mix of enterprises. An earlier example discussed a farm record system data for dairy enterprise information. The analyst used auxiliary census and university data for weighting and examined census data and concluded that the mix of operations (based on size, technology, and geographic location) within the data system did not mirror the actual population of dairy enterprises about which the CAR estimates were produced. Each subgroup was represented but not proportionately. The analyst was able to use the secondary sources of data to adjust the statistical inference of estimates produced using primary data. The end result was a better product.

Documentation of data sources is a critical part of the overall documentation of CAR estimation process. If multiple data sources are used, multiple sources must be documented. This documentation should outline the role of each different data source in the estimation process. For each data source, the documentation should discuss the target population of the data set and any reliability issues relevant to the estimation process.

FUTURE POSSIBILITIES

The section “Data Sources” identified the three major alternative sources of data used in developing CAR estimates. Two of those sources, probability surveys and farm record systems, deserve further consideration together. Both involve collecting data directly from a group of farm operators. Both have very distinct advantages and disadvantages as a source of CAR data. This section discusses the potential for integrating those two data collection alternatives to produce a new source of data that builds upon the strengths of its parents.

As highlighted earlier, probability surveys, such as the ARMS, can provide statistical inference for a broad array of potential target populations within the farm sector, and can provide statistical measures of precision. However, the ARMS suffers from several types of nonsampling errors, does not produce longitudinal data, and has confidentiality restrictions imposed upon the release of data records. Farm record systems contain data that are longitudinal and generally regarded as highly accurate. However, they are difficult to use for multistate analysis because procedures across states are often incompatible. Furthermore, farms included in these data sets are not representative of the farm sector, or even sizeable subsets of that sector.

Probability surveys and farm record systems each have strengths in areas where the other has relative disadvantages. Herein lies an opportunity for future possibilities: the integration of a USDA cost of production survey with a network of university farm record systems. The target population would be medium to larger farm operators engaged in specified farm enterprises. The purpose of the integrated system would be to exploit the advantages of both data sources, to combine resources and produce a single data set for access by both USDA and university analysts, and to reduce the multiple demands for financial information on the farm operator. A description of such an integrated system is outlined below.

Procedures For an Integrated USDA and University Farm Record System

1. USDA identifies long-term data components needed for federal programs. University specialists review these data components for consistency with underlying production practices, agricultural structures, and economic principles.

2. University specialists incorporate USDA data needs into farm record systems as part of a core set of variables. Individual systems include those additional data elements needed for state programs. The core will remain consistent across systems. USDA reviews system implementation for consistency.

3. USDA selects a probability sample of farm operators.

4. USDA contacts farm operators in sample, explains program and obtains cooperation. USDA subsidizes enrollment fees for participants from probability sample.

5. University specialist enrolls selected farm operators into farm record system.

6. University specialist compiles information from system and furnishes farm record data to USDA.

7. USDA provides comprehensive farm record data on a confidential basis to university specialists.

8. University specialists furnish farm business analysis and consultation to farm operators.

A project of this type would require close cooperation among universities, and between those universities and USDA. The resulting integrated system could provide a new standard of excellence in CAR data. The following table clearly displays the likely advantages. The future potential of this approach is worthy of serious consideration.

TABLE 12.1 Characteristics of Alternative Data Collection Systems

Characteristics of Data	Current USDA Survey	Current University Farm Record System	Integrated USDA and University System
Based on Probability Sampling	YES	NO	YES
Consistent Procedures Used Across States	YES	NO	YES
Data Accuracy * Accurate reporting * Detailed Information * Close local scrutiny	MODERATE	HIGH	HIGH
Longitudinal Data	NO	YES	YES
Cost	HIGH	HIGH	HIGH
Data Availability	Limited†	Varies By State	YES‡

† Currently limited to in-house use by USDA and specific collaborators on site at USDA in Washington, D.C.
‡ Could be made available to university specialists on an as-needed basis.

APPENDIX 12A

OVERVIEW OF STATISTICAL SAMPLING TECHNIQUES

This appendix provides an overview of various statistical sampling techniques, geared toward the analyst who either plans to collect primary data for CAR estimation, or who needs to understand the sample design that produced an existing data set. The overview provides limited details. Those readers desiring a more thorough discussion of these and other techniques should consult with one of the following texts that provide good introductory discussions of sampling: *Introduction to Survey Sampling* by Graham Kalton and *A Sampler on Sampling* by Bill Williams. More technical discussions are provided in *Sampling Techniques* by William Cochran and in *Sample Survey Methods and Theory* by Hansen et al.

Simple Random Sampling

This is the most basic type of statistical sampling. It involves selecting units from the population with equal probabilities, similar to drawing balls from an urn or names from a hat. An example in the context of CAR estimation: obtain a list of producers within a county, number each producer on the list and use a "random number table" to select which ones are included in the sample. Simple random sampling can be done "with" or "without" replacement, depending on whether a selected unit also can be selected on a subsequent draw (ball is replaced in the urn after it is drawn). In the CAR context, one would want to select producers "without" replacement. In practice, simple random sampling is seldom used because there are many more efficient, albeit more complex, alternatives.

Systematic Sampling

This is a variation of simple random sampling, which involves listing the population units into a random or purposeful ordering, selecting a "random start," and then selecting every "n^{th}" unit in sequence. Following the CAR example from above, if you wanted to select one out of every ten producers on the list, you would select a random number between 1 and 10 for the random start. If the random start was "3," your sample would include the producers numbered 3, 13, 23, etc.

The ordering is an important part of this sampling procedure. Random ordering will produce results similar to simple random sampling. Purposeful ordering will produce somewhat different results. For example, ordering the population by size or geographic location would force diversity within the sample. If the producer list in the CAR example had axillary information that showed acres operated, you could order the list by acres to ensure that a cross section of large, medium, and small producers would be in the sample.

Systematic sampling helps assure that the sample will adequately represent the diversity within the population. However, it does have its negative side. Undetected cycles in the ordering could lead to serious sampling bias. Systematic sample is generally operationally easier to do than simple random sampling.

Stratified Sampling

Stratified sampling forces diversity into the sample and at the same time reduces the variances of the estimates produced. The essence of this technique is to subdivide the population such that the resulting groups are fairly homogeneous regarding the attribute being measured. The groups are called strata. Each stratum is sampled separately, using a technique such as simple random sampling or systematic sampling. Stratified sampling is very common in survey sampling.

Extending the CAR example from above, you could stratify the names on the list into small, medium, and large producer groups based on acres operated. You would then sample each of the three groups, or strata, separately. The procedure would force diversity into the sample, similar to systematic sampling. Unlike that example, during estimation the variance is calculated within each stratum and then added across strata. If the operations within each group are fairly homogenous relative to enterprise CARs, the within- strata variances will be small, making the overall variances of the CAR estimates lower than under systematic sampling.

Single and Multistage Cluster Sampling

Cluster sampling is another technique in which the analyst divides the population into groups for more effective sampling. The goal is to reduce costs or save time. In stratified sampling, one samples within *each* group independently. In cluster sampling, one first selects a sample of these groups or clusters, and then, if necessary, selects units within the cluster in a second stage of sampling. Cluster samples are very effective when one does not have a complete listing of the population from which to sample.

Extending the CAR example from above, suppose the analyst wants to sample producers in the entire state instead of just a single county. It would be too costly and time-consuming to develop a list of producers in each county. The analyst first makes a list of all counties within the state. These counties represent clusters of producers. The analyst then randomly selects a sample of counties using one of the methods previously discussed. This is the first stage of the sample, and the counties are referred to as “primary sampling units.” For the *selected counties only*, a list of producers is developed. Each of these lists is sampled separately to obtain a sample of producers. This is the second stage of sampling, and the producers are referred to as “secondary sampling units.” Thus the analyst has obtained a representative sample of producers throughout the state without having to build a list of all producers.

Probability Proportional to Size Sampling

This technique is generally applied with cluster sampling. If the clusters are not the same size, and they most often are not, one may not want to give each the same chance of selection. Instead one would want to give the larger clusters, with the most secondary units, a larger chance of selection.

Extending the CAR example above, one would want to sample counties proportional to the number of producers within each county. Thus a county with twice as many producers would be twice as likely to be in the sample. However, since you do not have that information available, you could use census data to sample counties proportional to total production (of the commodity of interest).

APPENDIX 12B

GUIDELINES FOR ROUNDING CAR ESTIMATES

Consistent rounding of data and estimates expedites the analyst's comprehension of numerical data and provides an indication of the precision of the estimates. Two guidelines for rounding should be followed in publishing CAR estimates.

First, published estimates should never display greater precision than the least precise input datum. For example, the average price of a chemical input should not be published to the nearest "tenth of a cent" when the input data producing that average was received in whole cents. Likewise, the average price of purchased livestock should not be published to the nearest cent when the input prices were received to the nearest dollar.

Second, estimates should be rounded based on their overall magnitude. The following tables should be used as guidelines. The first table is for production numbers, and the second for dollars.

IF ESTIMATE FALLS IN THIS RANGE	ROUND TO NEAREST:
1 - 99	1
100 - 999	10
1,000 - 9,999	100
10,000 - 99,999	500
100,000 - 999,999	1,000
1,000,000 +	10,000

IF ESTIMATE FALLS IN THIS RANGE	ROUND TO NEAREST:
< $1.00	TENTH OF CENT
$1.00 - $9.99	CENT
$10.00 - $99.99	TEN CENTS
$100.00 - $999.99	DOLLAR
$1,000.00 - $9,999.99	TEN DOLLARS
$10,000.00 +	HUNDRED DOLLARS